

REFINED HEINZ-KATO-LÖWNER INEQUALITIES

STEFAN STEINERBERGER

ABSTRACT. A version of the Cauchy-Schwarz inequality in operator theory is the following: for any two symmetric, positive definite matrices $A, B \in \mathbb{R}^{n \times n}$ and arbitrary $X \in \mathbb{R}^{n \times n}$

$$\|AXB\| \leq \|A^2X\|^{\frac{1}{2}} \|XB^2\|^{\frac{1}{2}}.$$

This inequality is classical and equivalent to the celebrated Heinz-Löwner, Heinz-Kato and Cordes inequalities. We characterize cases of equality: in particular, after factoring out the symmetry coming from multiplication with scalars $\|A^2X\| = 1 = \|XB^2\|$, the case of equality requires that A and B have a common eigenvalue $\lambda_i = \mu_j$. We also derive improved estimates and show that if either $\lambda_i \lambda_j = \mu_k^2$ or $\lambda_i^2 = \mu_j \mu_k$ does not have a solution, i.e. if $d > 0$ where

$$d = \min_{1 \leq i, j, k \leq n} \{|\log \lambda_i + \log \lambda_j - 2 \log \mu_k| : \lambda_i, \lambda_j \in \sigma(A), \mu_k \in \sigma(B)\} \\ + \min_{1 \leq i, j, k \leq n} \{|2 \log \lambda_i - \log \mu_j - \log \mu_k| : \lambda_i \in \sigma(A), \mu_j, \mu_k \in \sigma(B)\},$$

then there is an improved inequality

$$\|AXB\| \leq (1 - c_{n,d}) \|A^2X\|^{\frac{1}{2}} \|XB^2\|^{\frac{1}{2}}$$

for some $c_{n,d} > 0$ that only depends only on n and d . We obtain similar results for the McIntosh inequality and the Cordes inequality and expect the method to have many further applications.

1. INTRODUCTION AND STATEMENT OF RESULTS

1.1. **Introduction.** We study the inequality

$$\|AXB\| \leq \|A^2X\|^{\frac{1}{2}} \|XB^2\|^{\frac{1}{2}},$$

where $A, B \in \mathbb{R}^{n \times n}$ are symmetric and positive definite, $X \in \mathbb{R}^{n \times n}$ is completely arbitrary and the norm is given by the operator norm

$$\|A\| = \sup_{\|x\|=1} \|Ax\|.$$

Fujii & Furuta [11] have shown that the special case

$$\|PQP\| \leq \|P^2Q\|^{\frac{1}{2}} \|QP^2\|^{\frac{1}{2}}$$

(P, Q both symmetric and positive definite) is equivalent to the following inequalities:

- **Löwner-Heinz inequality.** (Löwner [19], 1934), (Heinz [15], 1951). If $A \geq B \geq 0$, then

$$A^\alpha \geq B^\alpha \quad \text{for all } 0 \leq \alpha \leq 1$$

- **Heinz-Kato inequality.** (Heinz [15], 1951), (Kato [17], 1952). If A, B are positive operators such that $\|Tx\| \leq \|Ax\|$ and $\|T^*y\| \leq \|By\|$ for all $x, y \in H$, then

$$|\langle Tx, y \rangle| \leq \|A^\alpha x\| \|B^{1-\alpha} y\| \quad \text{for all } 0 \leq \alpha \leq 1$$

- **Cordes inequality.** (Cordes [9], 1987). For all symmetric and positive-definite A, B and all $0 \leq s \leq 1$

$$\|A^s B^s\| \leq \|AB\|^s.$$

2010 *Mathematics Subject Classification.* 15A58 (primary), 47A30 and 47A63 (secondary).

Key words and phrases. Cauchy-Schwarz inequality, Löwner-Heinz inequality, Heinz-Kato inequality, Cordes inequality, McIntosh inequality.

Classical work inspired by the original paper of Heinz [15] include a 1953 paper of Dixmier [10], a 1955 paper of Heinz [16] and a 1960 paper of Cordes [8]. There are now hundreds of papers concerned with variations of these inequalities, for a first introduction we refer to Bhatia & Kittaneh [3, 4, 5], Furuta [12], the books of Furuta [13] and Zhan [22] and references therein. A natural generalization of the inequality described above is due to A. McIntosh [20].

- **McIntosh inequality.** (McIntosh [20], 1979). Let $A, B \in \mathbb{R}^{n \times n}$ be symmetric and positive definite, let $X \in \mathbb{R}^{n \times n}$ be arbitrary and let $0 < r < 1$. Then

$$\|A^r X B^{1-r}\| \leq \|AX\|^r \|XB\|^{1-r}.$$

This inequality is easily seen to contain the previous inequality as a special case (relabeling $A \rightarrow A^2, B \rightarrow B^2$ and $r = 1/2$) and, in particular, it implies all previously mentioned results. Versions and variants in Hilbert space have been given by Bhatia & Davis [2] and Kittaneh [18].

1.2. Characterization of Equality. Despite a lot of activity surrounding these inequalities, the cases of equality are not known. If we consider diagonal matrices, it is obvious that equality will only result in conditions on some of the rows and cannot control the remaining entries (except that they have to be sufficiently small). More generally, if equality occurs then it is generically stable under small perturbation of those subspaces which play no role in the computation of the first singular vector and equality can therefore only impose restrictions on some of the subspaces. It is instructive to consider the special case of diagonal matrices

$$A = \text{diag}(\lambda_1, \dots, \lambda_n), \quad X = \text{diag}(x_1, \dots, x_n) \quad \text{and} \quad B = \text{diag}(\mu_1, \dots, \mu_n).$$

The natural renormalization $\|AX\| = 1 = \|XB\|$ boils down to

$$\max_{1 \leq i \leq n} |\lambda_i x_i| = 1 = \max_{1 \leq i \leq n} |x_i \mu_i|$$

under which the inequality simplifies to

$$\max_{1 \leq i \leq n} \lambda_i^r |x_i| \mu_i^{1-r} \leq 1.$$

The case of equality clearly requires the existence of a $1 \leq i \leq n$ for which $1/\lambda_i = |x_i| = 1/\mu_i$ (in which case a suitable vector v would be given by having its only nonzero entry at the i -th coordinate). We observe that this implies that A and B have to have a common eigenvalue and that if $\|AXBv\| = 1$, then XBv is also an eigenvector of A . We show this to be a general description of the case of equality. As for notation, the spectral decomposition of A and B will be written as

$$Av = \sum_{k=1}^n \lambda_k \langle a_k, v \rangle a_k \quad \text{and} \quad Bv = \sum_{k=1}^n \mu_k \langle b_k, v \rangle b_k.$$

Throughout the rest of the paper we use $\pi_\mu : \mathbb{R}^n \rightarrow \mathbb{R}^n$ to denote the spectral projector onto the space spanned by the eigenvalues of B associated to the eigenvalue μ .

Theorem 1 (Equality in the McIntosh inequality.). *Let A, X, B be as above and scaled in such a way that $\|AX\| = 1 = \|XB\|$, let $v \in \mathbb{R}^n$ be normalized $\|v\| = 1$ and assume*

$$\|A^r X B^{1-r} v\| = \|AX\|^r \|XB\|^{1-r} = 1.$$

If $X\pi_\mu v \neq 0$, then $X\pi_\mu v$ is an eigenvector of A with eigenvalue μ . In particular, A and B have at least one common eigenvalue.

This shows that the cases of equality are fairly restrictive: any nontrivial interaction between A and B that is created by X has to have matching eigenvalues – to a certain extent this means that equality in the special case of diagonal matrices already paints a fairly complete picture of what can happen. As expected, this characterization of equality does not impose any restriction on parts of spectra that play no role in determining the value of the operator norm.

The proof uses intuition coming from interpolation of analytic operators. A careful analysis of a suitably taylored holomorphic function shows that (under $\|AX\| = 1 = \|XB\|$), cases of equality require

$$\forall t \in \mathbb{R} \quad \|A^{1+it}XB^{-it}v\|^r = 1.$$

An instructive (albeit slightly inaccurate) visualization the underlying geometry is as follows: for any vector v the matrix B^{-it} introduces oscillation separately in every eigenspace with a speed that is proportional to the eigenvalue. These rotations are then being mapped to $AXB^{-it}v$ which is then subjected to rotations induced by A^{it} . The final result is a certain lack of cancellation which requires the rotations induced by A^{it} to cancel those of $AXB^{-it}v$. This is only possible if their speed is matched which requires common eigenvalues.

1.3. Improved estimates. Our approach also gives a way of deriving improved estimates: we show that as soon as one of the equations

$$\lambda_i \lambda_j = \mu_k^2 \quad \text{and} \quad \lambda_i^2 = \mu_j \mu_k$$

does not have a solution, there is a quantitative improvement of the McIntosh inequality that only depends on how close the equation is to being solvable. Assuming, as we always do, the normalization $\|AX\| = 1 = \|XB\|$, we introduce a notion of distance between $\sigma(A)$ and $\sigma(B)$

$$\begin{aligned} d = & \min_{1 \leq i, j, k \leq n} \{ |\log \lambda_i + \log \lambda_j - 2 \log \mu_k| : \lambda_i, \lambda_j \in \sigma(A), \mu_k \in \sigma(B) \} \\ & + \min_{1 \leq i, j, k \leq n} \{ |2 \log \lambda_i - \log \mu_j - \log \mu_k| : \lambda_i \in \sigma(A), \mu_j, \mu_k \in \sigma(B) \}. \end{aligned}$$

Note that Theorem 1 implies that in the case of equality, there actually exists a solution to $\lambda_i^2 = \mu_j^2$ and we always have $d = 0$. However, as soon as $d > 0$, we can give a quantitative improvement that only depends on n, r and d .

Theorem 2 (Refined McIntosh inequality). *For every $0 < r < 1$ and every $d > 0$ there exists $c_{n,r,d} > 0$ such that*

$$\|A^r X B^{1-r}\| \leq (1 - c_{n,r,d}) \|AX\|^r \|XB\|^{1-r}.$$

The constant $c_{n,r,d}$ arises the solution of a problem in approximation theory which we explain at the end of the paper. We are not aware of this particular problem ever having been treated before and prove only a basic result; improved results would imply better quantitative control on the size of $c_{n,r,d}$. The arguments in this paper only imply

$$c_{n,r,d} \geq c_r \exp\left(-\frac{1}{c_r} \frac{\sqrt{n}}{d}\right) \quad \text{for some } c_r > 0 \text{ depending on } r.$$

The statement is easily illustrated: for any $a, b > 0$

$$\begin{aligned} a^r &= \left\| \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}^r \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & b \end{pmatrix}^{1-r} \right\| \\ &\leq \left\| \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\|^r \left\| \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & b \end{pmatrix} \right\|^{1-r} = a^r \max(1, b^{1-r}) \end{aligned}$$

We see that the inequality is sharp for $0 \leq b \leq 1$ but that an improvement becomes possible as soon as $b > 1$ and the scale of the possible improvement depends only on b and r . We have

$$\frac{\sigma(A)}{\|AX\|^r} = \{0, 1\} \quad \frac{\sigma(B)}{\|XB\|^{1-r}} = \left\{ \frac{1}{\max(1, b^{1-r})}, \frac{b}{\max(1, b^{1-r})} \right\}$$

and see that $d > 0$ if and only if $b > 1$ and $0 < r < 1$.

1.4. Cordes inequality. We emphasize that our approach is not limited to the McIntosh inequality. We illustrate this by treating the Cordes inequality [9] and obtain the same type of result using essentially the same argument: recall that the Cordes inequality states that for all symmetric and positive-definite $A, B \in \mathbb{R}^{n \times n}$ and all $0 \leq s \leq 1$

$$\|A^s B^s\| \leq \|AB\|^s.$$

This inequality has been of the continued interest (see, for example, applications related to the geometry of C^* -algebras [1, 6, 7]), however, cases of equality are not known. If $s \in \{0, 1\}$, the inequality is trivially true and no information about A and B can be deduced. We write the spectral decomposition of A and B as

$$Av = \sum_{k=1}^n \lambda_k \langle a_k, v \rangle a_k \quad \text{and} \quad Bv = \sum_{k=1}^n \mu_k \langle b_k, v \rangle b_k.$$

Under a suitable normalization $\|AB\| = 1$, the case of equality requires $\lambda_i \mu_j = 1$ to have a solution. Moreover, the only way for the inequality to be attained for a vector v is 'diagonal' action: all eigenvectors b_k of B for which $\langle v, b_k \rangle \neq 0$ are also eigenvectors of A with the inverse eigenvalue, i.e. $A^s B^s b_k = A^s \mu_k^s b_k = \mu_k^s A^s b_k = b_k$. We also obtain an improvement as soon as either

$$\frac{1}{\lambda_i} \frac{1}{\lambda_j} = \mu_k^2 \quad \text{or} \quad \frac{1}{\lambda_i^2} = \mu_j \mu_k \quad \text{does not have a solution.}$$

More precisely, assuming again the rescaling $\|AB\| = 1$, we introduce a parameter $d^* \geq 0$ via

$$\begin{aligned} d^* &= \min_{1 \leq i, j, k \leq n} \{|\log \lambda_i + \log \lambda_j + 2 \log \mu_k| : \lambda_i, \lambda_j \in \sigma(A), \mu_k \in \sigma(B)\} \\ &+ \min_{1 \leq i, j, k \leq n} \{|2 \log \lambda_i + \log \mu_j + \log \mu_k| : \lambda_i \in \sigma(A), \mu_j, \mu_k \in \sigma(B)\}. \end{aligned}$$

Theorem 3 (Refined Cordes inequality). *Let $0 < s < 1$, $\|AB\| = 1$, let $v \in \mathbb{R}^n$ be normalized $\|v\| = 1$ and assume*

$$\|A^s B^s v\| = \|AB\|^s = 1.$$

If $\pi_\mu v \neq 0$, then $\pi_\mu v$ is an eigenvector of A with eigenvalue $1/\mu$. If $d^ > 0$, then*

$$\|A^s B^s\| \leq (1 - c_{n,s,d^*}) \|AB\|^s.$$

The precise behavior of c_{n,s,d^*} is determined by the same problem in approximation theory that already determines $c_{n,s,d}$ in Theorem 2. The proof of Theorem 3 is merely a minor variation of the previous arguments used in Theorem 1 and Theorem 2 – indeed, we believe the argument to be applicable to a large class of inequalities.

2. PROOF OF THEOREM 1

2.1. A holomorphic function. Given a symmetric, positive definite matrix $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and a vector $v \in \mathbb{R}^n$, we write its spectral decomposition as

$$Av = \sum_{k=1}^n \lambda_k \langle v, a_k \rangle a_k.$$

This has a natural extension to complex powers: for $z \in \mathbb{C}$, we define

$$A^z v = \sum_{k=1}^n \lambda_k^z \langle v, a_k \rangle a_k,$$

which, for a fixed vector v , is merely a vector whose entries are sums of complex exponentials. Note that, for $\lambda \geq 0$,

$$\lambda^{it} = \cos((\log \lambda) t) + i \sin((\log \lambda) t) = e^{i(\log \lambda)t}.$$

In particular, we have that A^{it} is a unitary matrix since

$$\begin{aligned} \left\| \sum_{k=1}^n \lambda_k^{it} \langle v, a_k \rangle a_k \right\|^2 &= \left\| \operatorname{Re} \sum_{k=1}^n \lambda_k^{it} \langle v, a_k \rangle a_k + \operatorname{Im} \sum_{k=1}^n \lambda_k^{it} \langle v, a_k \rangle a_k \right\|^2 \\ &= \left\| \sum_{k=1}^n \cos((\log \lambda_k) t) \langle v, a_k \rangle a_k \right\|^2 + \left\| \sum_{k=1}^n \sin((\log \lambda_k) t) \langle v, a_k \rangle a_k \right\|^2 \\ &= \sum_{k=1}^n [\cos^2((\log \lambda_k) t) + \sin^2((\log \lambda_k) t)] |a_k|^2 = \sum_{k=1}^n |a_k|^2 = \|v\|^2. \end{aligned}$$

Since the desired inequality is invariant under multiplication by scalars, we can assume without loss of generality that

$$\|AX\| = 1 = \|XB\|.$$

We will henceforth assume v to be an arbitrary but fixed vector with $\|v\| = 1$. The proof consists of a detailed analysis of the behavior of the vector-valued map

$$z \rightarrow A^{1-z} X B^z v \quad \text{on} \quad \{z \in \mathbb{C} : 0 \leq \operatorname{Re} z \leq 1\}.$$

We are mainly interested in the size of the arising vectors, which would motivate taking the complex inner product $\langle w, w \rangle_{\mathbb{C}} := w \cdot \overline{w}$, however, this quantity is not holomorphic. Instead, we consider, for fixed $\|v\| = 1$, the complex-valued map

$$z \rightarrow \langle A^{1-z} X B^z v, A^{1-z} X B^z v \rangle_{\mathbb{R}} \quad \text{on} \quad \{z \in \mathbb{C} : 0 \leq \operatorname{Re} z \leq 1\}, \quad \text{where} \quad \langle v, w \rangle_{\mathbb{R}} = \sum_{k=1}^n v_i w_i.$$

We observe that every single entry in the vector

$$A^{1-z} X B^z v \quad \text{is holomorphic and thus so is} \quad \langle A^{1-z} X B^z v, A^{1-z} X B^z v \rangle_{\mathbb{R}}.$$

Of course, the Cauchy-Schwarz inequality is still valid since

$$|\langle v, w \rangle_{\mathbb{R}}| = \left| \sum_{i=1}^n v_i w_i \right| \leq \sum_{i=1}^n |v_i w_i| \leq \left(\sum_{i=1}^n |v_i|^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^n |w_i|^2 \right)^{\frac{1}{2}} = \|v\| \|w\|.$$

We will now consider $z = 0 + it$ for $t \in \mathbb{R}$. Using Cauchy-Schwarz, we see that

$$|\langle A^{1-it} X B^{it} v, A^{1-it} X B^{it} v \rangle| \leq \|A^{1-it} X B^{it} v\|^2.$$

A^{-it} and B^{it} are unitary matrices and $\|v\| = 1$, therefore

$$\|A^{1-it} X B^{it} v\| = \|A X B^{it} v\| \leq \|A X\| \|B^{it} v\| = 1.$$

The same reasoning implies that for $z = 1 + it$

$$\|A^{it} X B^{1-it} v\| = \|X B^{1-it} v\| \leq \|X B\| \|B^{-it} v\| = 1.$$

Using the trivial estimate

$$\begin{aligned} \|A^{1-z} X B^z v\| &\leq \|A^{1-z}\| \|X\| \|B^z v\| \leq \|A\|^{1-\operatorname{Re} z} \|X\| \|B\|^{\operatorname{Re} z} \\ &\leq \max(\|A\|, 1) \|X\| \max(\|B\|, 1) \end{aligned}$$

it is easy to see that

$$|\langle A^{1-z} X B^z v, A^{1-z} X B^z v \rangle_{\mathbb{R}}| \leq 1 \quad \text{on} \quad \{z \in \mathbb{C} : 0 \leq \operatorname{Re} z \leq 1\}.$$

Let us now assume that indeed

$$\|A^r X B^{1-r} v\| = \|A X\|^r \|X B\|^{1-r} = 1.$$

This means that for some vector $\|v\| = 1$, we have $\|A^r X B^{1-r} v\| = 1$. Note that for this particular choice $z = 1 - r$ every single entry is real-valued and therefore

$$\langle A^r X B^{1-r} v, A^r X B^{1-r} v \rangle_{\mathbb{R}} = 1.$$

However, we are dealing with a holomorphic function that is uniformly bounded on the entire domain and has boundary values of size at most 1: the maximum principle then implies that

$$\langle A^{1-z} X B^z v, A^{1-z} X B^z v \rangle_{\mathbb{R}} = 1 \quad \text{for all } \{z \in \mathbb{C} : 0 \leq \operatorname{Re} z \leq 1\}.$$

The next step is to analyze the implications of this equation for $z = 0 + it$ (one could perform an equivalent analysis on $z = 1 + it$), where the equation simply becomes

$$\langle A^{1-it} X B^{it} v, A^{1-it} X B^{it} v \rangle_{\mathbb{R}} = 1 \quad \text{for all } t \in \mathbb{R}.$$

We now compute these matrices:

$$B^{it} v = \sum_{l=1}^n \mu_l^{it} \langle v, b_l \rangle b_l \quad \text{and thus} \quad X B^{it} v = \sum_{l=1}^n \mu_l^{it} \langle v, b_l \rangle X b_l.$$

Furthermore

$$A^{1-it} X B^{it} v = \sum_{k=1}^n \lambda_k^{1-it} \langle X B^{it} v, a_k \rangle a_k$$

and therefore

$$\begin{aligned} \langle A^{1-it} X B^{it} v, A^{1-it} X B^{it} v \rangle_{\mathbb{R}} &= \left\langle \sum_{k=1}^n \lambda_k^{1-it} \langle X B^{it} v, a_k \rangle a_k, \sum_{k=1}^n \lambda_k^{1-it} \langle X B^{it} v, a_k \rangle a_k \right\rangle_{\mathbb{R}} \\ &= \sum_{k=1}^n \lambda_k^{2-2it} \langle X B^{it} v, a_k \rangle_{\mathbb{R}}^2 \\ &= \sum_{k=1}^n \lambda_k^{2-2it} \left\langle \sum_{l=1}^n \mu_l^{it} \langle v, b_l \rangle X b_l, a_k \right\rangle_{\mathbb{R}}^2 \\ &= \sum_{k=1}^n \lambda_k^{2-2it} \left(\sum_{l=1}^n \mu_l^{it} \langle v, b_l \rangle \langle X b_l, a_k \rangle \right)^2 = 1 \end{aligned}$$

2.2. Cases of Equality: Proof of Theorem 1.

Proof. The precise coefficients are not as important as the algebraic structure: writing

$$\alpha_k = \log(\lambda_k), \quad \beta_l = \log(\mu_l), \quad \text{and} \quad c_{k,l} = \lambda_k \langle v, b_l \rangle \langle X b_l, a_k \rangle$$

allows to notationally simplify the equation to

$$\sum_{k=1}^n e^{-2\alpha_k it} \left(\sum_{l=1}^n c_{k,l} e^{\beta_l it} \right)^2 = 1.$$

Note that all coefficients are real-valued. The remainder of the proof is algebraic: the only way for an expression of this type to be identically 1 for all values of $t \in \mathbb{R}$ is for all terms (except constants) to cancel. We start with the assumption that every eigenvalue has multiplicity one and the eigenvalues are given by $\beta_1 < \dots < \beta_n$.

Fact. If $c_{k,l} \neq 0$, then $\alpha_k = \beta_l$. In particular, the equation $\alpha_k = \beta_l$ has a solution.

Consider the quantities

$$\underline{\sigma} = \min \{2\beta_l - 2\alpha_k : c_{k,l} \neq 0\} \quad \text{and} \quad \overline{\sigma} = \max \{2\beta_l - 2\alpha_k : c_{k,l} \neq 0\}.$$

These numbers give the smallest and largest occurring frequencies: an explicit expansion shows that the algebraic structure forces the leading coefficients to be sums of squares. This implies

that there is no form of cancellation and allows us to identify the smallest and largest occurring frequencies as $\underline{\sigma}$ and $\bar{\sigma}$:

$$\sum_{k=1}^n e^{-2\alpha_k it} \left(\sum_{l=1}^n c_{k,l} e^{\beta_l it} \right)^2 = e^{i\underline{\sigma}t} \left(\sum_{\substack{k,l=1 \\ 2\beta_l - 2\alpha_k = \underline{\sigma}}}^n c_{l,k}^2 \right) + \sum_j d_j e^{ie_j t} + e^{i\bar{\sigma}t} \left(\sum_{\substack{k,l=1 \\ 2\beta_l - 2\alpha_k = \bar{\sigma}}}^n c_{l,k}^2 \right)$$

where the d_j, e_j could be explicitly computed and the arising frequencies satisfy $\underline{\sigma} < e_j < \bar{\sigma}$. However, in order for this expression to be 1, we require

$$\underline{\sigma} = 0 = \bar{\sigma},$$

which was the desired statement. It remains to deal with the general case, where eigenvalues may have multiplicities. Let us assume there are exactly m distinct eigenvalues $\beta_1 < \dots < \beta_m$. Then we may write the equation as

$$\sum_{k=1}^n e^{-2\alpha_k it} \left(\sum_{l=1}^m \langle X\pi_{\beta_l} v, a_k \rangle e^{\beta_l it} \right)^2 = 1$$

and the same argument as before applies: either $\langle X\pi_{\beta_l} v, a_k \rangle = 0$ or $\alpha_k = \beta_l$. This means that if $X\pi_{\beta} v \neq 0$, then it has to be mapped to the eigenspace of A associated to the eigenvalue β . \square

3. PROOF OF THEOREM 2

3.1. Two simple Lemmata. We need a fairly simple result stating that trigonometric functions oscillate to a certain extent on a fixed interval if their frequencies are bounded away from 0.

Lemma 1. *For every $\delta > 0$ and $n \in \mathbb{N}$, any $d_1 < d_2 < \dots < d_n$ with $|d_j| \geq \eta$ and any $c_j \in \mathbb{R}$*

$$\int_{-2\sqrt{n}/\delta}^{2\sqrt{n}/\delta} \sum_{j=1}^n c_j e^{id_j t} dt \leq \frac{1}{2} \frac{4\sqrt{n}}{\delta} \left\| \sum_{j=1}^n c_j e^{id_j t} \right\|_{L^\infty(\mathbb{R})}$$

Proof. It is easy to see that

$$\begin{aligned} \left| \frac{\delta}{4\sqrt{n}} \int_{-2\sqrt{n}/\delta}^{2\sqrt{n}/\delta} \sum_{j=1}^n c_j e^{id_j t} dt \right| &\leq \sum_{j=1}^n |c_j| \left| \frac{\delta}{4\sqrt{n}} \int_{-2\sqrt{n}/\delta}^{2\sqrt{n}/\delta} e^{id_j t} dt \right| \\ &= \sum_{j=1}^n |c_j| \sin\left(\frac{2\sqrt{n}d_j}{\delta}\right) \frac{\delta}{2\sqrt{n}d_j} \\ &\leq \sup_{t \geq 2\sqrt{n}} \left| \frac{\sin t}{t} \right| \sum_{j=1}^n |c_j| \leq \frac{1}{2\sqrt{n}} \sum_{j=1}^n |c_j|. \end{aligned}$$

It remains to bound the L^∞ -norm. We first observe that

$$2T \left\| \sum_{j=1}^n c_j e^{id_j t} \right\|_{L^\infty(\mathbb{R})}^2 \geq \int_{-T}^T \left| \sum_{j=1}^n c_j e^{id_j t} \right|^2 dt = \int_{-T}^T \sum_{j=1}^n c_j^2 + \sum_{\substack{j,k=1 \\ j \neq k}}^n c_j c_k e^{i(d_j - d_k)t} dt$$

and dividing by T and letting $T \rightarrow \infty$ shows with Cauchy-Schwarz

$$\frac{1}{2\sqrt{n}} \sum_{j=1}^n |c_j| \leq \frac{1}{2} \left(\sum_{j=1}^n c_j^2 \right)^{\frac{1}{2}} \leq \frac{1}{2} \left\| \sum_{j=1}^n c_j e^{id_j t} \right\|_{L^\infty}.$$

\square

It could be desirable to prove a version of this inequality for shorter intervals of size $\sim \delta^{-1}$. Such a result would make our improvement $c_{n,r,d}$ independent of n (although, of course, the size of the matrix still enters implicitly through d).

Lemma 2. Let $f : [0, \infty] \rightarrow \mathbb{R}^+$ be monotonically decreasing. If $g \in L^\infty(\mathbb{R})$ satisfies

$$\int_0^x g(z)dz \leq \frac{x}{2} \|g\|_{L^\infty(\mathbb{R})}, \quad \text{then} \quad \int_0^x f(z)g(z)dz \leq \|g\|_{L^\infty(\mathbb{R})} \left(\int_0^{3x/4} f(z)dz - \int_{3x/4}^x f(z)dz \right).$$

Proof. Before we sketch the argument, we find it helpful to display the extremal function g that maximizes $\int_0^x f(z)g(z)dx$ (and does so for every nonnegative, monotonically decreasing f).

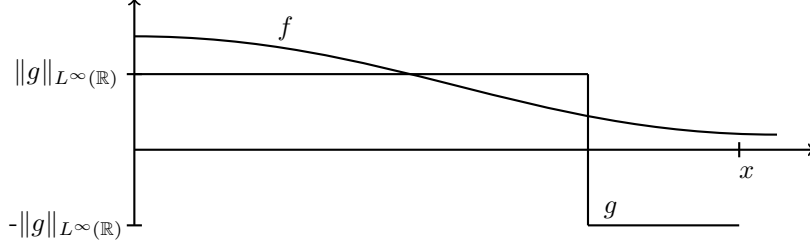


FIGURE 1. Maximizing the integral over $f \cdot g$ subject to constraints.

The proof uses the classical Hardy-Littlewood rearrangement inequality [14], which implies that

$$\int_0^x f(z)g(z)dz \leq \int_0^x f^*(z)g^*(z)dz,$$

where f^* is the monotonically decreasing rearrangement of f on $[0, x]$. Since f is monotonically decreasing, we have $f^* = f$. Consider now

$$h(z) = \begin{cases} \|g\|_{L^\infty(\mathbb{R})} & \text{if } 0 \leq z \leq 3x/4 \\ -\|g\|_{L^\infty(\mathbb{R})} & \text{if } 3x/4 \leq z \leq x. \end{cases}$$

It now suffices to show that

$$\int_0^x f(z)h(z)dz - \int_0^x f(z)g^*(z)dz = \int_0^x f(z)(h(z) - g^*(z))dz \geq 0.$$

This follows immediately from the fact that

$$\int_0^x h(z) - g^*(z)dz \geq 0, \quad \text{the definition of } h \text{ and the monotonicity of } f.$$

□

We will use this Lemma on a symmetric interval $[-x, x] \subset \mathbb{R}$ with f satisfying $f(x) = f(-x)$ in which case the result states that if

$$\int_{-x}^x g(z)dz \leq x\|g\|_{L^\infty(\mathbb{R})}, \quad \text{then} \quad \int_{-x}^x f(z)g(z)dz \leq 2\|g\|_{L^\infty(\mathbb{R})} \left(\int_0^{3x/4} f(z)dz - \int_{3x/4}^x f(z)dz \right).$$

3.2. Proof of Theorem 2. We can now prove the refined inequality.

Proof. Fix $0 < r < 1$ as well as the matrices A, X and B . We can assume again

$$\|AX\| = \|XB\| = 1$$

and $d > 0$, where

$$\begin{aligned} d = & \min_{1 \leq i, j, k \leq n} \{ |\log \lambda_i + \log \lambda_j - 2 \log \mu_k| : \lambda_i, \lambda_j \in \sigma(A), \mu_k \in \sigma(B) \} \\ & + \min_{1 \leq i, j, k \leq n} \{ |2 \log \lambda_i - \log \mu_j - \log \mu_k| : \lambda_i \in \sigma(A), \mu_j, \mu_k \in \sigma(B) \}. \end{aligned}$$

It remains to show that for all normalized vector $\|v\| = 1$

$$\|A^r X B^{1-r} v\| \leq 1 - c_{n,r,d}$$

for some constant $c_{n,r,d} > 0$ that only depends on these parameters. Consider the holomorphic map

$$z \rightarrow \langle A^{1-z} X B^z v, A^{1-z} X B^z v \rangle_{\mathbb{R}} \quad \text{on} \quad \{z \in \mathbb{C} : 0 \leq \operatorname{Re} z \leq 1\}.$$

It suffices to prove that this function at $z = 1 - r + 0i$ is at most $1 - c_{n,r,d}$ for all $\|v\| = 1$. Since the function is holomorphic, we can rewrite its value at a point as the integral over the associated Poisson kernel P paired with the boundary values of the function

$$\langle A^{1-z} X B^z v, A^{1-z} X B^z v \rangle_{\mathbb{R}} = \int_{\partial\{z \in \mathbb{C} : 0 \leq \operatorname{Re} z \leq 1\}} P(z) \langle A^{1-z} X B^z v, A^{1-z} X B^z v \rangle_{\mathbb{R}} dz.$$

We will now restrict our attention to one of the two lines of the boundary depending on which of the two parts of the separation measure d is bigger: we assume w.l.o.g.

$$\min \{ |2 \log \lambda_i - \log \mu_j - \log \mu_k| : \lambda_i \in \sigma(A), \mu_j, \mu_k \in \sigma(B) \} \geq \frac{d}{2}$$

and focus on $z = 0 + it$. If this was not the case, we would get the inequality for the other expression and focus on $z = 1 + it$. Since

$$|\langle A^{1-z} X B^z v, A^{1-z} X B^{2z} v \rangle_{\mathbb{R}}| \leq 1 \quad \text{on the entire strip,}$$

it suffices to prove for some $c_{n,r,d}^{(2)} > 0$

$$\int_{z=0+it} P(z) \langle A^{1-z} X B^z v, A^{1-z} X B^{2z} v \rangle_{\mathbb{R}} dz \leq \left(1 - c_{n,r,d}^{(2)}\right) \int_{z=0+it} P(z) dz.$$

An expansion of the function shows that we can write it as

$$\langle A^{1-z} X B^z v, A^{1-z} X B^{2z} v \rangle_{\mathbb{R}} = \sum_{k=1}^n \lambda_k^{2-2it} \left(\sum_{l=1}^n \mu_l^{it} \langle v, b_l \rangle \langle X b_l, a_k \rangle \right)^2 = \sum_j c_j e^{id_j t},$$

where

$$c_j \in \mathbb{R} \quad \text{and} \quad d_j \in \{\log \mu_i + \log \mu_j - 2 \log \lambda_k : 1 \leq i, j, k \leq n\} \subset \mathbb{R}$$

and therefore $|d_j| \geq d/2$. We can invoke Lemma 1 and guarantee that on an interval $[-\ell, \ell] \subset i\mathbb{R}$ centered around the origin (where ℓ only depends on n, r, d)

$$\int_{-i\ell}^{i\ell} \langle A^{1-z} X B^z v, A^{1-z} X B^{2z} v \rangle_{\mathbb{R}} dz \leq \ell.$$

The Poisson kernel $P(z)$ is symmetric around the x -axis, has its local maximum at $z = 0$, is monotonically decaying away from that maximum and is nonnegative everywhere. Lemma 2 applied to $f = P$ and g being the trigonometric function the implies

$$\int_{-i\ell}^{i\ell} P(z) \langle A^{1-z} X B^z v, A^{1-z} X B^{2z} v \rangle_{\mathbb{R}} dz \leq \int_{-\frac{3i\ell}{4}}^{\frac{3i\ell}{4}} P(z) dz - 2 \int_{\frac{3i\ell}{4}}^{i\ell} P(z) dz$$

and this gives the desired result. \square

4. PROOF OF THEOREM 3

Proof. The argument is a straightforward adaption of the existing argument. We normalize $\|AB\| = 1$ and consider the complex-valued map

$$z \rightarrow \langle A^z B^z v, A^z B^z v \rangle \quad \text{on} \quad \{z \in \mathbb{C} : 0 \leq \operatorname{Re} z \leq 1\}.$$

Trivially

$$|\langle A^{it} B^{it} v, A^{it} B^{it} v \rangle| \leq 1$$

because all matrices are unitary and we have

$$|\langle A^{1+it} B^{1+it} v, A^{1+it} B^{1+it} v \rangle| \leq 1$$

because of the normalization $\|AB\| = 1$. If indeed

$$\|A^s B^s v\| = 1 \quad \text{for some } 0 < s < 1,$$

then the maximum principle implies that

$$\langle A^z B^z v, A^z B^z v \rangle = 1 \quad \text{on} \quad \{z \in \mathbb{C} : 0 \leq \operatorname{Re} z \leq 1\}.$$

We analyze this equation for $z = 0 + it$ for $t \in \mathbb{R}$. Clearly,

$$\|A^{it} B^{it} v\|^2 = \sum_{k=1}^n \lambda_k^{2it} \left(\sum_{l=1}^n \mu_l^{it} \langle v, b_l \rangle \langle b_l, a_k \rangle \right)^2 = 1.$$

This falls within the general setup and the result follows from a repetition of previous arguments. \square

5. SOME REMARKS

5.1. Approximation Theory. This section states the approximation problem for which refined results would imply improved quantitative results about the behavior of the constants $c_{n,r,d}$ and c_{n,r,d^*} (and would also extend to other stability results obtained via the method outlined above). Consider the strip $\{z \in \mathbb{C} : 0 \leq \operatorname{Re} z \leq \pi\}$, where the constant π is chosen to simplify notation. A solution of the Dirichlet problem on that strip is given as follows (see e.g. [21]): define

$$P(x, y) = \frac{\sin x}{\cosh y - \cos x}.$$

This function is nonnegative, decays exponentially and satisfies

$$\int_{-\infty}^{\infty} P(x, y) dy = 2(\pi - x) \quad \text{for all } 0 < x < 2\pi.$$

A solution of the Dirichlet problem on the strip is given by

$$u(x, y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(x, t - y) u(0, t) dt + \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\pi - x, t - y) u(\pi, t) dt.$$

Fix now $0 < r < 1$, define the function $g(y) = P(\pi r, y)/(2\pi(1 - r))$ (this normalization turns g into a probability distribution) and the function class

$$\mathcal{F} = \left\{ \sum_{k=1}^n e^{-2a_k it} \left(\sum_{l=1}^n c_{k,l} e^{b_l it} \right)^2 : a_k, b_l, c_{k,l} \in \mathbb{R} \right\},$$

which contains trigonometric functions with real coefficients that have a special algebraic structure. \mathcal{F} contains the constant function 1 and therefore

$$\sup_{\substack{f \in \mathcal{F} \\ \|f\|_{L^\infty(\mathbb{R})} \leq 1}} \int_{-\infty}^{\infty} f(y) g(y) dy = 1.$$

The question is now under which conditions on a_k and b_k it is possible to obtain improved results. Lemma 2 shows that if we define $\mathcal{H} \subset \mathcal{F}$ via

$$\mathcal{H} = \left\{ f \in \mathcal{F} : \min_{1 \leq i, j, k \leq n} |b_i + b_j - 2a_k| \geq \delta \right\},$$

then

$$\sup_{\substack{f \in \mathcal{H} \\ \|f\|_{L^\infty(\mathbb{R})} \leq 1}} \int_{-\infty}^{\infty} f(y) g(y) dy = 1 - c_{n,r,d} < 1 = \int_{-\infty}^{\infty} g(y) dy.$$

One natural question is whether the same result holds in the larger space $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ given via

$$\mathcal{G} = \left\{ f \in \mathcal{F} : \min_{1 \leq i, j \leq n} |a_i - b_j| \geq \delta \right\},$$

which would then yield stability version that only depend on the minimal difference of elements in $\sigma(A)$ and $\sigma(B)$, which would naturally complement our characterization of equality.

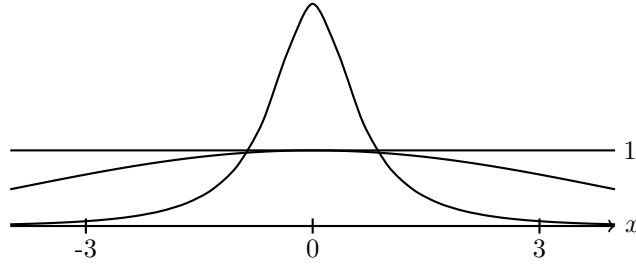


FIGURE 2. An example for $r = 0.6$ and the function $\cos^2(0.2x) \in \mathcal{F}$.

5.2. Simple Corollaries. A consequence of our equality characterization is that McIntosh inequalities with different indices are tightly linked.

Corollary. Let $0 < r < 1$, $\|AX\| = 1 = \|XB\|$, let $v \in \mathbb{R}^n$ be normalized $\|v\| = 1$ and assume

$$\|A^r X B^{1-r} v\| = 1.$$

Then for all $0 < s < 1$

$$\|A^s X B^{1-s} v\| = 1.$$

The crucial ingredient in our argument was that whenever the equation is attained, then

$$\langle A^{1-z} X B^z v, A^{1-z} X B^z v \rangle_{\mathbb{R}} = 1 \quad \text{for all } \{z \in \mathbb{C} : 0 \leq \operatorname{Re} z \leq 1\}.$$

This immediately implies the statement. The same argument also holds for the Cordes inequality.

Corollary. Let $0 < r < 1$, $\|AB\| = 1$, let $v \in \mathbb{R}^n$ be normalized $\|v\| = 1$ and assume

$$\|A^r B^r v\| = 1.$$

Then for all $0 < s < 1$

$$\|A^s B^s v\| = 1.$$

Another simple consequence of the proof is a slightly refined McIntosh inequality.

Corollary. Let $A, B \in \mathbb{R}^{n \times n}$ be symmetric and positive definite and let $X \in \mathbb{R}^{n \times n}$ be arbitrary. Then, for every $v \in \mathbb{R}^n$,

$$\|A^r X B^{1-r} v\| \leq \sup_{t \in \mathbb{R}} \|A^{1+it} X B^{-it} v\|^r \cdot \sup_{t \in \mathbb{R}} \|A^{it} X B^{1-it} v\|^{1-r} \leq \|AX\|^r \|XB\|^{1-r}$$

This follows immediately from an application of the Hadamard three-line theorem.

Acknowledgement. This paper grew naturally out of a long series of enjoyable discussions with Raphy Coifman for which the author is very grateful.

REFERENCES

- [1] E. Andruchow, G. Corach and D. Stojanoff, Geometrical significance of Löwner-Heinz inequality. Proc. Amer. Math. Soc. 128 (2000), no. 4, 1031–1037.
- [2] R. Bhatia and C. Davis, A Cauchy-Schwartz inequality for operators with applications. Linear Algebra Appl. 223/224 (1995), 119–129.
- [3] R. Bhatia and F. Kittaneh, Notes on matrix arithmetic-geometric mean inequalities. Linear Algebra Appl. 308 (2000), no. 1-3, 203–211.
- [4] R. Bhatia and F. Kittaneh, Norm inequalities for positive operators. Lett. Math. Phys. 43 (1998), no. 3, 225–231.
- [5] R. Bhatia and F. Kittaneh, The matrix arithmetic-geometric mean inequality revisited. Linear Algebra Appl. 428 (2008), no. 8-9, 2177–2191.
- [6] G. Corach, Operator inequalities, geodesics and interpolation. Functional analysis and operator theory (Warsaw, 1992), 101–115, Banach Center Publ., 30, Polish Acad. Sci., Warsaw, 1994.
- [7] G. Corach, H. Porta and L. Recht, An operator inequality. Linear Algebra Appl. 142 (1990), 153–158.
- [8] H. Cordes, A matrix inequality. Proc. Amer. Math. Soc. 11 (1960) 206–210.
- [9] H.O. Cordes, Spectral Theory of Linear Differential Operators and Comparison Algebras, London Mathematical Society Lecture Note Series, vol. 76, Cambridge University Press, Cambridge, 1987

- [10] J. Dixmier, Sur une inégalité de E. Heinz. Math. Ann. 126, (1953). 75–78.
- [11] M. Fujii and T. Furuta, Löwner-Heinz, Cordes and Heinz-Kato inequalities. Math. Japon. 38 (1993), no. 1, 73–78.
- [12] T. Furuta, Norm inequalities equivalent to Löwner-Heinz theorem. Rev. Math. Phys. 1 (1989), no. 1, 135–137.
- [13] T. Furuta, Invitation to linear operators. From matrices to bounded linear operators on a Hilbert space. Taylor & Francis, Ltd., London, 2001.
- [14] G. E. Hardy, J. E. Littlewood, and G. Polya. Inequalities. Second edition, Cambridge University Press, London and New York, 1952.
- [15] E. Heinz, Beiträge zur Störungstheorie der Spektralzerlegung, Math. Ann. , 123 (1951) pp. 415–438.
- [16] E. Heinz, On an inequality for linear operators in a Hilbert space. Report of an international conference on operator theory and group representations, Arden House, Harriman, N. Y., 1955, pp. 27–29. Publ. 387. National Academy of Sciences-National Research Council, Washington, D. C., 1955.
- [17] T. Kato, Notes on some inequalities for linear operators, Math. Ann. 125 (1952) pp. 208–212.
- [18] F. Kittaneh, Norm inequalities for fractional powers of positive operators. Lett. Math. Phys. 27 (1993), no. 4, 279–285.
- [19] K. Löwner, Über monotone Matrixfunktionen, Math. Z., 38 (1934) pp. 177–216.
- [20] A. McIntosh, Heinz inequalities and perturbation of spectral families, Macquarie Mathematics Reports, Macquarie University, 1979.
- [21] D. Widder, Functions harmonic in a strip. Proc. Amer. Math. Soc. 12 (1961) 67–72.
- [22] X. Zhan, Matrix inequalities. Lecture Notes in Mathematics, 1790. Springer-Verlag, Berlin, 2002.

(Stefan Steinerberger) DEPARTMENT OF MATHEMATICS, YALE UNIVERSITY, NEW HAVEN, CT 06510, USA
E-mail address: `stefan.steinerberger@yale.edu`